Capstone Project - The Battle of Neighborhoods

Opening an "Indian Restaurant" in Montreal, Canada.

By,

Akash Janakiraman

INTRODUCTION

This is a capstone project for IBM Data Science Professional Certificate. In this project, we imagine a scenario for a concept that there is a person who wants to open an Indian Restaurants in Montreal- Quebec, Canada. The person wants to start his business in a place where he can maximize his profits. One way to go ahead with the problem is starting to analyze places where many people visit so that many customers will visit the restaurant. One more approach to the problem is analyzing the city and opening a restaurant in an area where there is less number of Indian restaurants or competitors. Here we try to use the second method to obtain the best results.

BUSINESS PROBLEM

The objective of this capstone project is to find the most suitable location for an entrepreneur to open a new Indian Restaurant in Montreal, Canada. By using data science methods and tools along with machine learning algorithms such as K-means clustering, this project aims to provide solutions to answer the business question: In Montreal, if an entrepreneur wants to open an Indian Restaurant, where should they consider opening it?

TARGET AUDIENCE

The entrepreneur wants to find the ideal location to open an authentic Indian restaurant. The majority of his customers will be people from the Asian community and tourists from abroad.

DATA

To solve this problem, we will need below data:

● List of neighbourhoods in Montreal, Canada
● Latitude and Longitude of these neighbourhoods
● Venue data related to Indian restaurants. This will help us find neighbourhoods that are more suitable to open an Indian Restaurant.

EXTRACTING THE DATA

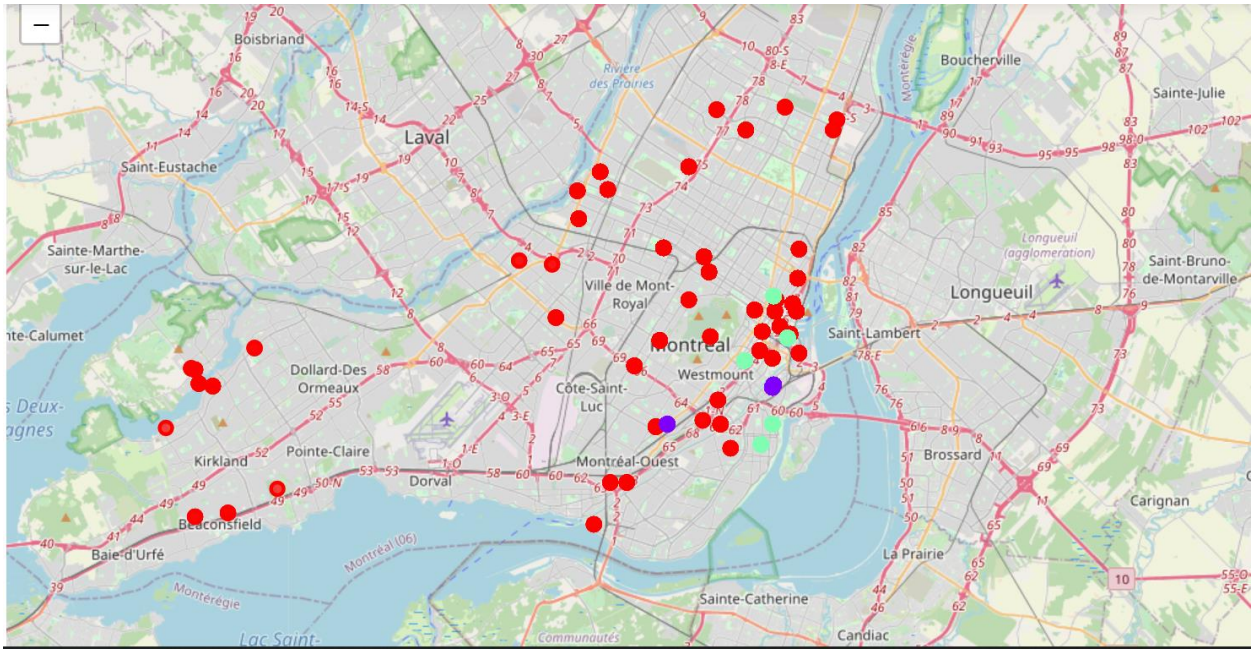
● The scrapping of Montreal neighbourhoods data via Wikipedia.
● I am getting latitude and Longitude data of these neighbourhoods via Geocoder package.
● I am using Foursquare API to get venue data related to these neighbourhoods.

METHODOLOGY

First, I got the list of neighbourhoods in Montreal, Canada. This was done by extracting the list of neighbourhoods from Wikipedia: https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Montreal
I did the web scraping by utilizing pandas HTML table scraping method with Beautiful soup library as it is easier and more convenient to pull data directly from a web page into the data frame.
However, we get only the list of neighbourhoods names in Montreal. I obtained their coordinates using Foursquare API to pull the list of venues near these neighbourhoods. To get the location coordinates, I used the Geocoder Package. After gathering these coordinates, I visualized the map of Montreal using Folium package to verify whether these are correct coordinates. Next using the Foursquare API, I pulled the list of top 100 venues within 500 meters radius. I created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I got access to datasets like names, categories, latitude, and Longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyzed each neighbourhood by grouping the rows by neighbourhood and taking the mean on the frequency of occurrence of each venue category. This was done to prepare for the clustering algorithm to be implemented later.
Now, I categorized the data sets based on venue type as "Indian restaurants".
Lastly, I performed the clustering algorithm by using the k-means clustering algorithm. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms, and it is highly suitable for this project as well. I have

clustered the neighbourhoods in Montreal into 3 clusters based on their frequency of occurrence for "Indian food". Based on the results (the concentration of clusters), I have inferred the ideal location to open an Indian restaurant.

RESULT



The above map represents the results from k-means clustering algorithm and shows that we can categorize Montreal neighbourhoods into 3 clusters based on how many Indian restaurants are in each neighbourhood.

RECOMMENDATIONS

From the above results, we find that most of the "Indian Restaurants" are found in neighbourhoods classified under cluster 2. So, it won't be profitable to open a "Indian Restaurant" in the neighbourhoods present in cluster 2 as there is a lot of competition from other restaurants. Similarly, in neighbourhoods found in cluster 1 contains many "Indian Restaurants". Therefore, it's not a good idea to open an

"Indian Restaurant" in those neighbourhoods present in cluster 1. Therefore, the ideal location to start an "Indian restaurant" will be any of the neighbourhoods present in cluster 0 like "Old Montreal", "Saint-Henri" etc., as there are a minimum number of competitors present in those locations.
The above inference was obtained by carrying out neighbourhood analysis using K-Means clustering algorithm.